Arms Control in the Information Age: World-Wide Data Acquisition, Analysis, Storage and Access in Near-Realtime

Steven R. Bratt
International Data Centre
Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization
P.O. Box 1200
A-1400 Vienna, AUSTRIA
Steve.Bratt@ctbto.org
tel +43-1-26030-6143
fax +43-1-26030-5923

Abstract

The verification system being established to monitor the Comprehensive Nuclear-Test-Ban Treaty (CTBT) will include a network of 337 seismic, hydroacoustic, infrasound and radionuclide facilities, transmitting digital data to the International Data Centre (IDC) in Vienna, Austria over a global satellite communications system. The mission of the IDC is to process and analyze these data, and to provide equitable, timely and convenient access to the data, products and services to all Member States so as to support their compliance assessments under the Treaty. In effect, the Treaty defines a new brand of arms control verification for the information age. Enabling technologies include near-realtime global acquisition of data, public key cryptography, automated knowledge-based data fusion and decision support, mass storage systems, and user interfaces based on Internet and Web technologies.

During 2000, the IDC disseminated products based on data from about 90 seismic, hydroacoustic, infrasound and radionuclide sensor stations of the future monitoring network. The number of events in the reviewed seismo-acoustic bulletins ranged from 40 to 360 each day. On average, some 200 radionuclide spectra were processed and analysed each month. Users from 45 Member States received an average of close to 18,000 data and product deliveries per month from the IDC. As the verification system and the base of users grow, the IDC will continue to integrate state-of-the art information technology and scientific methods to insure that the potential of the verification system is realized in terms of improved automation, speed, quality, cost-effectiveness, service and overall credibility.

1 Introduction

After 51 years of nuclear testing – including the detonation of more than 2,000 nuclear explosions – the Comprehensive Nuclear-Test-Ban Treaty (CTBT) was open for signature at United Nations Headquarters in New York on 24 September 1996 [1]. The CTBT bans all nuclear weapons test explosions or any other nuclear explosion in all environments. Entry into force (EIF) of the Treaty will occur once a particular 44 nations that both participated in the negotiations and that host nuclear research or power reactors have signed the Treaty and have deposited their instruments of ratification with the United Nations.

To verify compliance, the Treaty requires a verification system [2] (Figure 1) composed of an International Monitoring System (IMS) of geophysical and environmental sensors located worldwide, sending digital data via a Global Communications Infrastructure (GCI) to an International Data Centre (IDC) for analysis, long-term archiving and distribution to Member States (countries that sign and then ratify the Treaty). The verification regime also includes other elements (not considered in here) such as special event analysis, confidence-building measures and on-site inspection to investigate events considered by Member States to be of concern. The purpose of the verification regime is to ensure a high probability of detection for any non-compliance with the Treaty, and thus to provide a strong deterrent to clandestine nuclear explosions. The combination of the CTBT verification regime with other means of global monitoring (such as national technical means, and scientific and hazard monitoring networks) offers a high probability that relevant events will be detected, located and identified [3, 4]. In other words, the deterrent effect of this verification regime will be, in no uncertain terms, formidable.

The challenge faced by the new CTBT verification regime is a more ambitious extension of that faced by nations who have, for decades, monitored the nuclear testing activities of the nuclear powers. The signature of a nuclear explosion must be detected, located and identified against a backdrop of "noise". This "noise" is composed of low levels of natural and cultural background, in addition to larger signatures of earthquakes, volcanic explosions, conventional explosions detonated for commercial and military purposes, leakage from nuclear facilities, sonic booms, meteorological phenomena, etc. Furthermore, there is an increasing number of "noise" phenomena as the size or energy of the phenomena of interest decreases. This makes it an increasing challenge to detect and identify increasingly smaller nuclear explosions.

The CTBT will depend more strongly upon information technologies for its effectiveness than any arms control treaty in history. The large data volumes, remote sensor locations, existence of a level of mutual concern among nations and the need for objectivity and cost-efficiency all motivate the application of information technologies. Near-realtime data acquisition, digital data authentication, satellite-based Internet communications, knowledge-based processing and decision support, mass storage systems, and Web-based data access systems are some of the information technologies that are already being brought to bear on this problem of international prominence.

The Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization and its Provisional Technical Secretariat have as their very focused objective the implementation of this verification regime and the preparation of the regime for EIF. The Provisional Technical Secretariat was established in Vienna, Austria in March 1997, and has made significant progress in achieving this ambitious objective.

This paper provides an overview of the CTBT verification system. The remote sensor network of the IMS and the global communication facilities are described first. Next, the IDC processing and analysis systems are reviewed. Finally, more detailed information on the data acquisition, mass storage and access systems of the IDC is given. The paper includes also summaries of the results of test operations to date, and discusses the challenges that lie ahead.

2 International Monitoring System

To use an anatomical analog, the IMS network hosts the "senses" of the CTBT verification system, and is composed of geophysical and environmental sensor stations spread across the globe (Figure 2). The IMS concept extends and integrates the methods employed by nations to monitor each other over the past several decades. Historically, long-range (from thousands of kilometers) seismological observations have been used to monitor, primarily, underground testing. Hydroacoustic methods have been used to monitor the oceans. Infrasound (sound waves in the 0.02 – 10 Hz range), and subsequently satellite methods, have been used to monitor the atmosphere. Atmospheric environmental sampling has been employed to detect radioactive aerosols and gasses, which can provide the best, unambiguous evidence for identifying nuclear explosions. In addition, there are potential synergies between the monitoring methods. For example, seismic stations could observe signals from explosions in the oceans and atmosphere, while infrasound could detect explosions underground and underwater especially when the explosions occur near the boundary between different environments. Nuclear explosions in the oceans or underground which are poorly contained could release radionuclides into the atmospheric in addition to hydroacoustic and/or seismic energy into the ocean and earth.

When completed, the IMS will include 50 primary seismic, 11 hydroacoustic and 60 infrasound stations transmitting digital data continuously at a total rate on the order of 5 to 10 Gbyte/day. Some 120 auxiliary seismic stations will provide segments of data upon request from the IDC. All of the hydroacoustic and infrasound stations, and about 35 of the seismic stations are composed of arrays of instruments, spatially distributed over apertures of kilometers. This allows for estimation of the direction of the incoming signals and, in many cases, enhancement of the signal-to-noise ratio. Eighty radionuclide stations will collect particulate samples on filter papers and return gamma ray spectra on a daily basis. Half of these stations will be equipped also to collect and analyze noble gas samples. Finally, 16 radionuclide laboratories will provide a more detailed analysis of samples collected at the radionuclide stations when certain agreed radioisotopes have been observed during routine analysis.

The technical and operational specifications for the IMS stations are quite rigorous. For example, the goal is for at least 98% of all data from primary seismic, hydroacoustic and infrasound stations to be available in a timely manner for processing and analysis at the IDC: a requirement made more challenging by the fact that many of these stations are located in remote areas and harsh environments. The stations must incorporate tamper protection systems. The IMS data must be digitally-signed near their digital source, so that receiving data centers can validate that the data have not been altered.

Many of the best existing stations have been selected to be upgraded and to become part of the IMS network. However, most of the stations must be installed from scratch, and promising new sites are being selected to host these new facilities. As of the end of

2000, the PTS had invested about 40% of the approximately $215 million required to implement the IMS. The first new and upgraded stations started providing data to the IDC in 2000. Close to 115 of the 321 seismic, hydroacoustic, infrasound and radionuclide stations met or substantially met the rigorous IMS specifications by the end of that year. Those in the "substantially met" category typically require upgrades to implement authentication and communications systems. Tens of new IMS facilities around the world are currently at some stage of the installation process.

3 Global Communications

A critical task within the CTBT verification system is to ensure the prompt, reliable and secure collection and redistribution of data. Exchange of data, products and services is enabled by an agreed set of communications formats and protocols, applied across the "nervous system" of the verification regime: the Global Communications Infrastructure (GCI). Data and products are exchanged throughout the verification regime using the Continuous Data (CD) format and the International Monitoring System (IMS) format.

The CD format is used to transmit primary seismic, hydroacoustic and infrasound data via the Transmission Control Protocol/Internet Protocol (TCP/IP). These data are transmitted twenty-four hours per day from the stations to the IDC. CD is also used to forward continuous data to those Member States that wish to receive it. The latest version, CD-1.1, has been enhanced to better support applications-layer error correction, command and control, state-of-health monitoring and integration of digital signatures. Signatures are computed using the Digital Signature Algorithm (DSA) [5] – a so-called "public key" cryptographic method – using a private key stored securely at the station and the data frames themselves. Receivers use a public key corresponding to the private key to authenticate or verify that the data received have not been tampered with.

The IMS format has its roots in the Automatic Data Request Manager (AutoDRM) system [6] that is employed at tens of geophysical data centers worldwide. It was expanded for CTBT purposes to support a larger variety of data and products. The latest version, IMS-2.0, is the principle format for transmission of requests from the IDC to auxiliary seismic stations for data segments, and for the return of the segmented timeseries. Radionuclide stations send their spectra to the IDC using this format as well. The underlying protocol is simply e-mail, and for large payloads, the File Transfer Protocol (FTP). Each IMS-2.0 e-mail contains a message in a documented, computer-readable format. Through the use of the S-MIME standard [7], the messages are accompanied by a certificate housing the sender's public key and the digital signature corresponding to the message. As with the CD transmissions, DSA is the algorithm used for securing the transmission of IMS data.

The GCI "nervous system" provides an integrated, cost-effective, high-availability communications service for the verification regime. The GCI is the first truly global (with sites on all continents and in all ocean regions) satellite communications network based on Very Small Aperture Terminal (VSAT) technology. The conceptual design is quite straightforward: IMS facilities and Member States in all but the near-polar regions of the world can exchange data starting from their local VSAT earthstation, via C-band transmissions through one of three geosynchronous satellites. The satellites route transmissions to one of three communications hubs on the ground, and then on to the IDC via terrestrial Frame Relay links, with ISDN back-up. In reality (Figure 3) the GCI also supplements the three global beams with two additional Ku-band satellites and hubs for more economical coverage of North America and Europe. Based on requests from the host nations, about one third of the IMS stations, may have their data routed through national communications nodes before they are routed into the GCI. An affordable solution for communicating with a handful of stations in the near-polar regions is still being sought. Overall, the GCI system must be cost-effective, operate with 99.5% availability and provide data within seconds from origin to final destination.

All of the planned hubs and terrestrial links of the $70 million (over 10 years) GCI are now operational. During 1999 the first IMS data started arriving at the IDC and the first data and products were distributed using the GCI. As of the end of 2000, a total of 44 VSATs had been installed at IMS facilities and at establishments of Member States authorized to access data and products. About 70 additional sites are in preparation.

About 30 seismic primary, 35 seismic auxiliary, four hydroacoustic, six infrasound and 15 radionuclide stations transmitted data into the IDC for processing and analysis during 2000. Some of these transmitted data via the GCI, and the remainder via nationally-furnished links used for testing purposes until the GCI links can be established. Experience indicates that the availability of data arriving through the GCI is significantly higher (~95%) as compared to the availability of data arriving through the other communications arrangements (~70%). Of course, the data availability at the IDC is a function not only of the GCI link availability (which has been near the required 99.5%) but also of the availability of the station and IDC infrastructures. Continued improvement of the IMS, IDC and communications systems, and commissioning of new VSAT’s at IMS stations, will act to increase the total availability.

4 International Data Centre

The necessity for an IDC emerged from years of international experimentation [8, 9, 10] demonstrating that effective verification of a multi-lateral nuclear-test-ban requires acquisition of large volumes of monitoring data, and the application of state-of-the-art science and technology to process, analyze, interpret, and store these data. Thus, to continue the anatomical analogy, the IDC, not unlike the “brain”, is responsible for synthesizing the IMS data. This includes the execution of an agreed set of tasks on behalf of all Member States; tasks which most nations, individually, would find too fiscally, computationally and/or scientifically intensive to execute on a global and continuous basis. The services of the IDC allow verification to become a global, cooperative activity, and not just the activity of the few nations that currently possess nuclear weapons or that have the necessary surveillance and computing assets. However, the IDC does not pass judgement to identify a particular event as a suspected nuclear test. This is because judgements regarding whether or not a particular event is of concern would most certainly depend upon not only the scientific parameters describing that event, but also upon the uncertainties in those parameters within the context of other information and political concern that one State might have about another.

The IDC is about halfway through an approximately five-year, seven-phase Progressive Commissioning Plan. The plan outlines a progressive hiring of qualified staff, procurement of computer hardware and commercial software, enhancement and testing of the monitoring applications software, and implementation of global communications capabilities. The plan also defines a progressive increase in the level of processing, analysis and service to the Member States, such that the IDC and States can prepare themselves for operations before EIF of the Treaty. With almost 100 on-board at this time, the staff of the IDC is planned to grow to over 130 in a manner that keeps pace with the growing IMS, and its growing list of developmental and operational responsibilities.

The vast majority of the software required to execute the monitoring functions outlined in the remainder of this section is being developed specifically for CTBT monitoring purposes. Much of this is being provided on a cost-free basis by the United States, through its prototype IDC in Arlington, Virginia [11]. Significant contributions are also being provided by other international monitoring experts and under contracts funded by the IDC. The estimated investment over the past five years in the monitoring software being provided by the United States is on the order of several tens of millions of dollars. The IDC expects to invest on the order of $2-4 million per year for the maintenance, improvement and documentation of the monitoring software.

Information technology plays a vital role in insuring that the mission of the IDC is executed in a manner that is as accurate, objective, rapid, cost-effective and reliable as possible, and that these all improve over time. In this rest of this section, the methods and technologies employed to support data acquisition, intelligent processing and analysis, data fusion and data access systems of the IDC are reviewed.

4.1 Data Acquisition and Quality Control

The IDC receives all data and validates the authenticity of these data using the accompanying digital signatures and a public key that corresponds to the private key used to sign the data at the station. The data are then placed on the "diskloop" of the Operational Data Management System (DMS) in a special format used for subsequent processing and analysis. Section 5 provides further detail regarding the IDC's data management architecture and concept of operation. Data quality is checked at various stages of acquisition, processing and analysis.

4.2 Intelligent Seismic, Hydroacoustic, Infrasound Processing and Analysis

Seismic, hydroacoustic and infrasound data are digital timeseries that are similar enough to be processed and analyzed [12] using the same processing and analysis “pipeline” (Figure 4). Small differences between the processing and analysis of data from one of these technologies versus another are generally related to specific processing parameters applied and to the particular algorithms executed.

Automated Station Processing. Within minutes after a segment of continuous primary seismic, hydroacoustic and infrasound data arrives at the IDC, it undergoes automated station processing. The output of this is a list of "signals" (where the signal level sufficiently exceeds the noise) detected within the timeseries from a single station, and a suite of attributes that describes each detected signal (e.g., arrival time, amplitude, frequency, direction, estimate of the travel path). Methods including digital filtering, beamforming (for array stations), signal detection, neural network techniques, Bayesian inferencing and rule-based analysis are applied during automated station processing.

Automated Network Processing. The list of signal detections and attributes from all stations in the network are next input into the automated network processing system. The data undergo three rounds of automated network processing, with each round integrating additional data. The output of each round is a Standard Event List (SEL). SEL1, SEL2 and SEL3 are compiled automatically within about two, six and 12 hours, respectively, after an event's occurrence. The automated network processing software is a grid-search, knowledge-driven system that associates signal detections with the hypothesis that those signals have propagated from an event in a particular cell within a grid of cells that covers the earth. This extensible approach supports the integration of source-, station- and path-specific information into each grid cell, including spatially-varying earth models and spatially- and temporally-varying ocean and atmospheric models. An iterative, least-squares inversion procedure uses the signal arrival times and other parameters to estimate the latitude, longitude, depth and origin time for each event [13]. Following the production of SEL1, discrete time periods of seismic data are automatically requested and received from the auxiliary seismic stations in a manner that seeks to improve the event solutions contained in the SEL1. The retrieved auxiliary data undergo automated station processing, with any signals detected used in the production SEL2 and SEL3, and in subsequent interactive analysis.

Interactive Analysis. Review of SEL3 by human analysts results in the compilation of a Reviewed Event Bulletin (REB) within two to three days after the occurrence of an event. A sophisticated graphical decision support and execution tool is used to review the raw timeseries data in conjunction with the automatic processing results within the SEL3, and to validate or correct, when necessary, the automated solutions.

Automated Characterization and Screening. Using the REB as a starting point, a number of automatic processes are initiated with the objective of estimating additional characteristics of each event [14]. The resulting product is the Standard Event Bulletin (SEB). Standard event screening criteria are next applied on an automated and objective basis to screen-out events considered, with high confidence, to be consistent with natural phenomena or non-nuclear, man-made phenomena. The resulting bulletin is the Standard Screened Event Bulletin (SSEB).

Seismic, Hydroacoustic and Infrasound Results. Figure 5 shows a map of the estimated event locations given in the REBs for 2000. An average of 52 events per day were located and provided within the daily REBs. The maximum daily number of events, 357, occurred on 16 November 2000, when two large main shocks occurred east of Papua New Guinea, followed by many tens of aftershocks. As the IMS network grows and improves, the improved detection capability of the verification system will allow for an even larger number of smaller events to be detected, located and characterized.

The synergy between the seismic, hydroacoustic and infrasound technologies is already being demonstrated, despite the small number of stations currently available in the latter two technologies. During 2000 only one test hydrophone station in the Pacific Ocean region transmitted data to the IDC, and then for just part of the year. Despite this, almost 20% of all events in the global REB, most of which are presumably earthquakes, are generating observable hydroacoustic phases. Also, despite the few infrasound stations providing data to the IDC during 2000, both seismic and infrasound signatures have been observed from presumed near-surface conventional explosions. These cross-technology observations will be of value in improving the ability to screen-out larger numbers of events [15].

4.3 Intelligent Radionuclide Processing and Analysis

Radionuclide processing and analyses proceed through a pipeline (Figure 4) that is similar at a high level to the one described above, but different in detail [16]. One difference is the time schedule. The propagation of radionuclides from the causative event to the sensors could take on the order of days to weeks, and the subsequent sample collection and measurement in the field takes up to two days for noble gas analysis and up to three for particulate analysis.

Automated Station Processing. Automated radionuclide processing at the IDC starts rapidly after each daily gamma-ray spectrum is received from an IMS radionuclide station. The processing represents an integrated extension of the standard approach to gamma-ray spectrometry. The processing sequence includes a preliminary peak search, an energy calibration update, a final peak search, efficiency calibration, a background subtraction and nuclide identification.

A field of regard (FOR) is defined as the geographic region from which the air mass sampled during the collection period is estimated to have been located at some time prior to collection. Fields of regard are estimated for every radionuclide sample, and for several time points (e.g., one day, two, days, etc.) in the past. This is done through an automated application of the HYSPLIT atmospheric transport software [17], with wind analysis data from the medium range forecast model operated by the National Centers for Enviornmental Prediction (NCEP) and received at the IDC over the Internet.

The Automatic Radionuclide Report (ARR) contains information on the collection of each sample; processing parameters; provisional peak detection and radionuclide identification; calibration information; the minimum detectable concentrations of key radionuclides and the FORs. Separate ARRs are provided for each air filter sample and each noble gas sample. The ARR for a particular sample is available to the Member States within minutes of the time each spectrum arrives at the IDC.

Interactive Analysis. Interactive analysis follows, with an expert analyst using graphics-based analysis software to validate or correct (if necessary) the peak picks and nuclide identifications determined by the automated system.

Automated Characterization and Screening. Following analyst review, the radioactivity measurements are automatically characterized by the number, type and relative amounts of nuclides present. Each spectrum is assigned a categorization "Level" based on the presence or absence of anomalous radionuclides from an agreed list of "relevant radionuclides" [18]. The results of the analyst review and automatic categorization processing for each sample are documented in a Reviewed Radionuclide Report (RRR).

Detailed Meteorological and Sample Analyses. Identification of anomalous levels of at least one relevant radionuclide triggers the start of a detailed atmospheric transport modeling operation. The IDC currently employs both the HYSPLIT and Omega [19] software to do this. The IDC could also cooperate with National and Regional Specialized Meteorological Centres of the World Meteorological Organization to support high-priority atmospheric transport analyses. Should multiple, anomalous, relevant nuclides be identified in a station's sample, the station's operator is instructed to dispatch the sample to the IMS radionuclide laboratories for further analysis.

A Standard Screened Radionuclide Event Bulletin (SSREB) is produced for those samples that contain one or more anomalous, relevant radionuclides. The SSREB includes all individual RRRs for all of individual spectra thought to be associated with the hypothesized source event; all atmospheric transport modeling results; and results of any laboratory analyses.

Radionuclide Results. The IDC processed and analyzed 2,601 specta during 2000. Following analyst review, there were 40 spectra indicating a single, relevant radionuclide and only indicating multiple, relevant radionuclides.

It is fortunate that there are not many instances during which there is a significant release of radionuclides. However, two types of sources of interest for validating the atmospheric transport models of the IDC are plumes from volcanoes and from large forest fires [20]. Figure 6 shows a satellite image of the plume from a large wildfire near San Diego, California in January 2001 as compared to the forward wind trajectories determined at the IDC using the Omega model. In this case, the comparison is encouraging. Much more work is needed to integrate appropriate transport models and meteorological data into IDC processing, to expand cooperation with global meteorological centers, and validate the processes employed.

4.4 Fusion, Review and Summary

The final stages of producing the standard products of the IDC are still under consideration. However, the following elements are being examined:

Four-Technology Data Fusion. The IDC would attempt, automatically and/or interactively, to cross-reference in space and time the event location and origin time estimates from the SSEBs as determined from the seismic, hydroacoustic and infrasound analysis, with the detailed FORs from the SSREBs estimated from the radionuclide and atmospheric transport analysis. The resulting Fused Event Bulletin (FEB) could also include context measures (geographic association of located events with features reported by Member States, such as active quarries or mines, earthquake faults, volcanic fields, power plants, or occurrences at or near these features) or other agreed characterization parameters.

Executive Summary. This product presents to the Member States on one electronic page an abstract of the most salient results for a given data day (Figure 7). Included are the number of seismic, hydroacoustic and infrasound events; the number of radionuclide events; the status of screening within both pipelines; any information on possible fused events; and statistics related to the state-of-health of the IMS and IDC systems. The report is a dynamic one that is updated, especially, as a result of the evolving processing and analysis of radionuclide and fusion over a period of two weeks or so.

4.5 System Monitoring

As a standard service, the status of all IMS stations, communications, product dissemination, and of the IDC systems is continuously monitored. This includes surveillance of the state-of-health, capability and security of these elements. Immediate notification is forwarded to the responsible parties (e.g., staff members or contractors) should the operational performance of any component of a system drop below the specified levels. An integrated system to support this requirement that has been offered by the Government of France is now being tested.

4.6 Data and Product Access

The objective of the IDC access services is to help Member States to find and retrieve those data of interest to them. The IDC automatically disseminates to each Member States the requested data and products without cost up to a certain total volume (currently 100 Mbyte per day per State). A Member State can request larger volumes of data, with possible cost implications. There are three methods for accessing data:

Subscriptions. Subscriptions can be established electronically that specify the types and amounts of data that a Member State wishes to receive (e.g., events from a particular region, signal detections at certain stations, continuous streams of timeseries data, etc.) and the frequency with which they wish to receive them (e.g., immediately as available, daily, etc.). Subscriptions are satisfied and forwarded (via the CD or IMS formats described in Section 3) automatically whenever the selections and frequency criteria are met, and continue until the subscription is cancelled or its specifications changed.

One-Time AutoDRM Requests. Member States can make one-time requests for specific data or products by sending a request message in IMS format via e-mail to the IDC. The response containing the requested data or product is automatically returned by the IDC via e-mail, unless its size exceeds an upper limit specified by the requestor. In this case, the requestor is automatically notified of the availability of the response via e-mail, and instructed how to retrieve the large file(s) via FTP.

Secure Web System. The IDC provides a secure Web site that supports interactive browsing, selecting and downloading of data, products and other information. The Web system also allows Member States to interactively browse all data and products, and to format requests for subscriptions or AutoDRM-type queries. Figure 7 shows the IDC products home page, which provides a view of the Executive Summary for a particular day. From here, Member States can review and download IMS data and IDC products available from this day or others. Products can be retrieved for a particular time period, or through a data-mining interface that supports searches for data or events meeting complex suites of spatial, temporal, geophysical and/or environmental criteria.

Data and Product Access Results. By the end of 2000, 318 authorized users from 47 Member States were authorized by their National Authorities to access IMS data and IDC products. An average of 18,000 products was shipped automatically to authorized users each month during the year. About 96% of all products were shipped within three minutes of the request time or scheduled subscription time. Most of the products are subscriptions for automatic SELs to provide early alerts for possible events of interest to Member States. An average of 1 Gbyte of raw data was accessed through AutoDRM each month. About 50 Gbyte per month of continuous, streaming IMS data were forwarded to subscribing Member States. A modest 400 sessions per month via the Web system were logged. These numbers will most certainly grow as the number of users grows and EIF of the Treaty approaches.

5 Computing and Data Management System

The IDC monitoring software described in Section 4 is a computationally-intense, data-centric system. In this section, the computing and data management infrastructure required to support the monitoring operations is reviewed in greater detail.

5.1 Computing and Security Infrastructure

The IDC is housed within 3,000 m^2 on two floors of the Vienna International Centre. The floors were re-constructed, largely under the funding of the Government of Austria, to support special space, security, power and environmental requirements. Over $11 million has been invested by the Preparatory Commission over the past three years to establish the IDC computing infrastructure (Figure 8), including the computing and storage servers, workstations, networking equipment, security systems and commercial off-the-shelf software. Requirements imposed by the software being transferred from the United States coupled with a look ahead to emerging technologies, led to the adoption of the following standards (as of the end of 2000): Sun servers and workstations, the Solaris 7 operating system, Oracle 8 RDBMS, and X-windows and Web for graphic user interfaces.

The network is divided into a number of subnetworks and services. Hosted within the IDC Operations Local Area Network (LAN) are the mission-critical functions of the IDC. The Development and the Test/Integration LANs conform to these same standards, so as to support repair, enhancement, integration and testing of new software. Other functional elements of the infrastructure support office automation, information security, state-of-health monitoring and networking with the GCI and Internet.

System integrity is maintained in the IDC using mechanisms for physical security (e.g., access doors and controls), system security (e.g., operating system features, software security packages, system password control, database management system security features and configuration management) and operational security (e.g., software design procedures, software implementation controls, start-up and shutdown procedures, system failure procedures, security reviews and accountability procedures). The IDC operates a Public Key Infrastructure to support the generation, exchange, management and revocation of digital certificates and keys for the public key cryptography scheme that is employed to sign and authenticate information exchanged within the verification system.

5.2 Data Management System Concept

The philosophical underpinnings of the IDC data management system (DMS) date to the early 1980s [21] when Lawrence Berkeley and Lincoln Laboratories, under funding by the Defense Advanced Research Projects Agency, designed an efficient DMS for the new Center for Seismic Studies (later to host the prototype IDC). It was conceived that the most effective DMS for the types of data and products required for monitoring nuclear testing would be a combination of an RDBMS and the Unix network file system (NFS). The voluminous raw timeseries (or other) data acquired from geophysical sensors are stored on the NFS. Meta or parametric data – including information about the station, channel, instrumentation, start time, sample rate, etc. – for each timeseries segment within each file are stored in the RDBMS, along with a pointer to the location of the timeseries on the NFS. The output of subsequent processing and analysis of the timeseries data is also stored in the RDBMS. Thus the RDBMS contains indices to the raw data, making it efficient to browse and retrieve the timeseries data by station, temporal and/or geophysical parameters. The RDBMS schema have evolved since the early 1980s, but the concept is much the same. The schema and variations on them are used widely in geophysical data centers throughout the world.

Another concept is that data arriving into the Operational DMS (from and to which the operational acquisition, processing and analysis systems input and output data) are migrated automatically, and as rapidly as practicable, to the larger Archive DMS. Data are automatically purged from the Operational DMS after between 10 and 20 days for the timeseries on the diskloop of the NFS, and after between four and six months for the parametric data in the RDBMS. These data migration policies act to stabilize the load on the operational DMS, and minimize the access to those data by non-operational entities (e.g., research staff, authorized users of the Member States, etc.)

5.3 Mass Storage System

The Mass Storage System (MSS) of the IDC was procured during 1999, and is now undergoing acceptance testing. The primary functions of the MSS are: (a) to receive and store raw timeseries data that have been migrated from the Operational DMS diskloop to the Archive DMS and (b) to service requests for these data as described in Section 4.6. Both of these functions are fully automated. The MSS acts like an integral part of the NFS within the Archive DMS, in that the IMS data stored on the MSS are indexed by pointers in the RDBMS. Some other types of data are also archived on the MSS, as will be described below. Hierarchical storage management (HSM) techniques are employed to optimize data storage and retrieval response times.

Due to the lack of practical experience, establishing the archival and retrieval specifications for procurement of the MSS was a challenge. Estimates of the volume of data to be archived on the MSS were the better understood of the two. The model was that all raw IMS data would be archived on the MSS at least twice: once in their native CD or IMS format (with their digital signatures, for later authentication) and once in the format used by the monitoring software during processing and analysis. At least one additional copy of the native format data would be archived off-site. Segments of timeseries data corresponding to located events (e.g. containing from a time just before the first signal detection to just after the expected time of the last theoretical signal) would be archived to allow rapid search and retrieval of event-based data. Other data to be archived include meteorological data that will be arriving on a daily basis, maps and images arriving periodically and some one-time archiving of historical data from the prototype IDC. Table 1 summarizes the estimated volumes of data to be archived on the MSS. In addition, the MSS may be used to store certain system back-ups.

The MSS will reside on the IDC Operations LAN, which is a 100 Mbit/s switched Ethernet network. Data will be archived to the MSS by one or more hosts on the this LAN. Over time, the operational procedures will be optimized so that only one host is archiving its data at a certain time, to avoid network congestion. Thus one requirement was that the I/O write bandwidth from the servers to the write unit(s) must exceed the bandwidth of one 100 Mbit/s network connection (12.5 Mbyte/s).

Archiving of data will be performed in steps. First, all available data segments from all stations for a particular period of time (tens of minutes to hours) will be packed into a standard TAR format file. Each file could contain tens to thousands of data segments, each indexed by a pointer in the RDBMS. Second, these files will be transferred from the Operations LAN diskloop to a disk cache on the MSS. Third, the files are copied automatically, using HSM principles, to tape on the MSS.

Delays in data acquisition (e.g., due to IMS, GCI or IDC disruptions) motivate two undesirable, possible remedies. One possibility is the IDC systems implement an unwelcome delay in the packing and archiving process to wait for possible late data. Alternatively, packing and archiving of data, even data with known gaps, could occur as rapidly as possible, with later-arriving data packed and archived in separate files. This would make later retrieval of contiguous segments of data slower, since more files might have to be read from tape to close all gaps. With increasingly reliable IMS, GCI and IDC systems, completeness of the data will be able to be ascertained, and the packed data files migrated to the MSS, at a much earlier time.

The retrieval of IMS data will be performed through requests for specific time-periods of data from specific stations and channels. Data still residing on the disk-cache will be transmitted rapidly, whereas data residing on tape will first be read from the library to the disk cache and then transmitted.

The model for retrieval of data from the MSS was more difficult to specify during the procurement phase, given that the IDC had no experience at that time with user request patterns. Even the experience of today is not close to what it will likely be, say, after EIF. Therefore, parameters such as the optimal size and location of files were estimated based partially on the experience of the prototype IDC, and partially on the maximum network bandwidth and I/O throughput provided by the IDC computer infrastructure. For example, it was assumed that at any one time, there could be up to five active retrieval requests, in addition to the IDC monitoring software activity. Thus it was specified that the I/O retrieval throughput from the read units to the server must exceed the bandwidth of seven active 100 Mbit/s network connections, which corresponds to a data transfer rate of approximately up to 87.5 Mbyte/s. These retrieval requests could occur simultaneously with the above-described archiving activity.

Based on these data archival and retrieval models, the IDC procured a solution centered on StorageTek technology (Figure 9). The heart of the new MSS includes a Sun Enterprise 4500 Server (two 450 MHz processors, 1 Gbyte memory), a StorageTek 9310 tape library (six StorageTek 9840 tape drives, 70 Tbyte of tape storage initially), Sun StorEdge A5200 disk cache (1.6 Tbyte cache, consisting of six RAID-5 260 Gbyte disk-arrays and one RAID-5 500 Gbyte disc-array for HSM RDBMS). Two redundant fiber-channels, three UWDSCSI and five 100BASE-T interfaces provide network connectivity. The system supports upgrade paths to ATM, Gigabit Ethernet and FDDI. The Application Storage Manager (ASM), Navisphere Agend, NAVICLI and ACLS software manage efficient archiving, retrieving and tape and disk control activities.

The peak data transfer rate of the offered disk-subsystem is 200 Mbyte/s, which exceeds the requirements. The average data transfer rate between the disk-subsystem and a write-unit is 10 Mbyte/s, and between the disk-subsystem and a read units is 6x10 Mbyte/s (both rates assuming hardware compression is disabled).

To assess the performance of the MSS versus the requirements, the IDC developed a benchmark data suite. Three typical retrieval requests were simulated:

- Small requests (<1 Mbyte) from one waveform file from N simultaneous users.
- Medium requests (> 1 and <10 Mbyte) from one or more waveform files from N simultaneous users
- Large requests (>10 and <32 Mbyte) from one or multiple waveform files from N simultaneous users

In total 1,612 sample requests were prepared, executed and analyzed during the procurement selection process at the vendor's facility. The data segments requested resided in files ranging from 34 Mbyte to 817 Mbyte in size, summing to a total volume of 26 Gbyte of raw timeseries data on the MSS. The test-configuration was run at the vendor's site, using an MSS with only one 9840 drive (in comparison with the six planned for the IDC). The suite of sample requests was completed within four hours, for an average response time of under 10 s per request. With multiple drives, the performance could be expected to improve significantly.

Final acceptance tests are now underway on the newly installed MSS within the IDC. During the course of operations, detailed data archival and retrieval statistics will be maintained to analyze the actual usage of the system. Adaptations to the structure and file architecture will be made as necessary.

6 Conclusions and Opportunities for the Future

The IDC has already established an advanced, integrated capability for data acquisition, storage, processing, analysis and dissemination to serve the needs of the CTBT's Member States. As the IMS network grows, as the GCI communications links bring increasing volumes of new data into Vienna, and as the IDC software and experience mature, the capability of the IDC and of the entire verification system will improve beyond the significant level currently being demonstrated as part of the Preparatory Commission's testing program. It would seem that this integrated application of information technologies employed within the CTBT verification regime presents an attractive model for future arms control and surveillance problems that are rich in data, are computationally intensive and have requirements for rapid analysis and data access.

The continued progressive commissioning process brings with it a number of interesting opportunities and challenges which, if met, will insure that the true performance of the verification system approaches its theoretical capabilities. Fortunately, the framers of the Treaty, foreseeing the scientific and technical challenges of the tasks they posed, envisioned that the technologies, methods and procedures would be progressively developed as experience and advances dictate. The major technical challenge is achieving the high target levels of availability for the IMS, the GCI and the IDC. To do this requires a coherent operation and maintenance concept for the overall verification; superior station, communications and information systems; and effective monitoring of the state-of-health of the systems, in order to sense degradation and to alert the people responsible for maintaining and repairing them.

There are a number of scientific opportunities as well. Methods to detect, locate and characterize small events near the threshold of detection of the network will continue to be improved. Fundamental research and development in the areas of the processing and analysis of hydroacoustic and infrasound data are, in a sense, allowing the state-of-the art in these areas to "catch-up" with the operational seismic and radionuclide processing and analysis. Work remains to improve event-screening methods, develop more efficient interactive analysis and decision support tools, and realize the potential for synergy and fusion of data between the four monitoring technologies. Improving the application of atmospheric transport analysis is also required to better address the four-technology fusion problem. Methods must be "tuned" and signal propagation models "calibrated" [22] to incorporate the best, available knowledge of potential sources; propagation paths through the earth, oceans and atmosphere; and station characteristics. All of this work must be integrated into a large, but modular, software system that can be maintained and improved over time.

Finally, there are a number of policy and procedural challenges that must be addressed. Many of these relate to classic trade-offs between the quality of products desired, the timeliness of the products and the costs of investments in personnel and system development to support the quality and timeliness levels sought.

There is a broad understanding within the Preparatory Commission of the potentially great value of the IMS data and IDC products to organizations that monitor disasters and provide relief (e.g., for earthquakes, tsunamis, volcanoes, radioactive fall-out, etc.) and to organizations that perform scientific research and development. So too it is understood that the CTBT could benefit from these organizations to help address the aforementioned challenges. The Preparatory Commission has made it a priority to complete a policy to support the provision of IMS data and IDC products to such organizations, so that these rich and unique assets can be of benefit to all who need them for verification, humanitarian and scientific endeavors.

Acknowledgements
The work described in this paper is the product of the efforts of scientists and engineers around the globe. The Conference on Disarmament, Group of Scientific Experts developed many of the concepts that are now being implemented in Vienna. The staff and contractors of the prototype IDC in Arlington, Virginia have played a significant role in the implementation and initial testing of much of the software currently running at the IDC. The staff of the IDC in Vienna have invested tireless hours to install, operationally test and assess the software, methods and procedures to support the requirements of the CTBT. I especially acknowledge the Director of the IDC, Rashad Kebeasy, for his leadership and support. Christian Korschan, John Coyne, Joseph Undercoffer and Joep Winkels contributed significant efforts in the procurement, installation and testing of the MSS, and its documentation in this paper. Lars-Erik De Geer, Petr Firbas Jan Wuster and Prajesh Bhakta provided thorough and helpful technical reviews of this manuscript. Finally, I thank Claire Taylor for her help with the figures and Jeanette Tirabassi and Silvy Thomas for their expert integration of the final paper.

Disclaimer: *The views expressed herein are those of the author, and do not necessarily reflect the views of the CTBTO Preparatory Commission.*

References

[1] United Nations General Assembly. *Comprehensive Nuclear-Test-Ban Treaty.* (New York, NY: United Nations Report, **A/50/1027,** 10 September 1996)

[2] W Hoffmann, R Kebeasy and P Firbas, “Introduction to the Verification Regime of the Comprehensive Nuclear-Test-Ban Treaty,” *Physics of the Earth and Planetary Interiors,* **113** (1999) 5–9.

[3] Independent Commission on the Verifiability of the CTBT. *Final Report*. (London: The Verification Research, Training and Information Centre, 2000)

[4] J M Shalikashvili. *Letter to the President and Report on the Findings and Recommendations Concerning the Comprehensive Nuclear Test Ban Treaty*. (Washington, DC: US Department of State Report, 2001)

[5] National Institute of Standards and Technology. *Digital Signature Standard (DSS)*. (Washington, DC: Federal Information Processing Standard Publication FIPS, **186-**1, U.S. Department of Commerce, December 30, 1996)

[6] U Kradolfer, "Automating the Exchange of Earthquake Information," *EOS Transactions, American Geophysical Union,* **74** (1993) 442.

[7] S Dusse et al. *S/MIME Version 2 Message Specification.* (Network Working Group, **RFC-2311**, 1998)

[8] S R Bratt. "An International Experiment in Rapid Exchange and Interpretation of Seismic Data from a Global Network (GSETT-2)," *EOS Trans. Amer. Geophys, Un.,* **73** (1992) 513, 520.

[9] F Ringdal, "GSETT-3: "A Test of an Experimental International Seismic Monitoring System," *Annali di Geophisica*, **XXXVII** (1994) 241-245.

[10] A Wuster, F Riviere, R Crusem, J-L Plantet, B Massinon and Y Caristan, "GSETT-3: Evaluation of the Detection and Location Capabilities of an Experimental Global Seismic Monitoring System," *Bull. Seis. Soc. Am*., **90** (2000) 166-186.

[11] J R Bowman, W Farrell and K Garcia. *Software of the Prototype IDC*. (San Diego, CA: Science Applications International Corporation Technical Report, **SAIC-96/ 1177,** 1996)

[12] Science Applications International Corporation. *IDC Processing of Seismic, Hydroacoustic, and Infrasonic Data.* (San Diego, CA: Science Applications International Corporation Technical Report, **SAIC-99/3023,** 1999)

[13] S R Bratt, and T C Bache, "Location Estimation Using Regional Array Data," *Bull. Seis. Soc. Am.* **78** (1988) 780-798.

[14] M D Fisk, D C Jepsen and J R Murphy, "Experimental Seismic Event-Screening Criteria at the Prototype International Data Centre," *Submitted to Pure and Applied Geophysics* (2001)

[15] G G Sorrells, E T Herrin and J L Bonner, "Construction of regional ground-truth databases using seismic and infrasound data," *Seism. Res. Lett.*, **68** (1997) 743-752.

[16] L R Mason, J D Bohner and D L Williams, "Global Radionuclide Monitoring in Near-Real Time for Verification of the Comprehensive Nuclear Test Ban Treaty," *Journal of Radioanalytical and Nuclear Chemistry*, **235** (1998) 65-69.

[17] R R Draxler and G D Hess, “An Overview of the HYSPLIT- 4 Modelling System for Trajectories, Dispersion and Deposition,” *Australian Meteorological Magazine*, **47** (1998) 295–308.

[18] De Geer, “CTBT Relevant get title from Lars-Erik,” *Submitted to Kerntechnick,* (2001).

[19] D P Bacon, N Nash’at Ahmad, Z Boybeyi, J D Thomas, M S Hall, C S Pius, R Lee, A Sarma and M T Turner, “A Dynamically Adapting Weather and Disperion Model,” *The Operational Multiscale Environment Model with Grid Adaptivity, Monthly Weather Review*, **128** (2000) 2044-2076.

[20] G Wotawa and M Trainer, "The Influence of Forest Fires on Pollutant Concentrations in the United States," *Science*, **288** (2000) 324-328.

[21] J Berger, R G North, R C Goff and M A Tiberio, “A Seismological Data Base Management System,” *Bull. Seism. Soc. Am*, **74** (1984) 1849–1862.

[22] P Firbas. *Location Calibration based on 3-D Modeling Steps towards the calibration of the global CTBT Network*, *in Advances in Seismic Event Location.* C Thurber and N Rabinowitz – editors, (Dordrecht, Netherlands: Kluwer Academic Publishers, 2000)

Table 1. Estimated Volumes of Data to be Archived on the IDC Mass Storage System

Data type	Compressed (before transfer?)	Note	Daily Volume (Mbyte)	Monthly Volume (Gbyte)	First Year Volume (Tbyte)	Five Year Volume (Tbyte)
Primary Seismic / IMS Raw Data	Yes	1	6172.28	191.34	2.30	11.48
Hydroacoustic / IMS Raw Data	Yes	1	2742.67	85.02	1.02	5.10
Infrasound/ IMS Raw Data	Yes	1	1366.18	42.35	0.51	2.54
Radionuclide/ IMS Raw Data	Yes	1	150	4.65	0.06	0.28
Auxiliary Seismic/ IMS Raw Data	Yes	1	424.53	13.16	0.16	0.79
Primary Seismic/ Processing Format	No	2,3	9351.94	289.91	3.48	17.39
Hydroacoustic/ Processing Format	No	2,3	3746.48	116.14	1.39	6.97
Infrasound/ Processing Format	No	2,3	2069.97	64.17	0.77	3.85
Radionuclide/ Processing Format	No	2,3	150	4.65	0.06	0.28
Auxiliary Seismic/ Processing format	No	2,3	643.23	19.94	0.24	1.20
Segment archive	No	2,3	2160.45	66.97	0.8	4.02
Meteorological data	No	1,3	7168	222.21	2.67	13.33
Map and imagery data	No	1,3	521.91	15.8	0.19	0.95
One time copy A from prototype IDC	1	3.4	Not applicable	Not applicable	1	1
One time copy B from prototype IDC	2	3.4	Not applicable	Not applicable	7	7
Total			**36667.64**	**1136.31**	**21.65**	**76.18**

*Notes: 1. Data read rarely. 2. Data read often. 3. Size of files can vary by an order of magnitude.

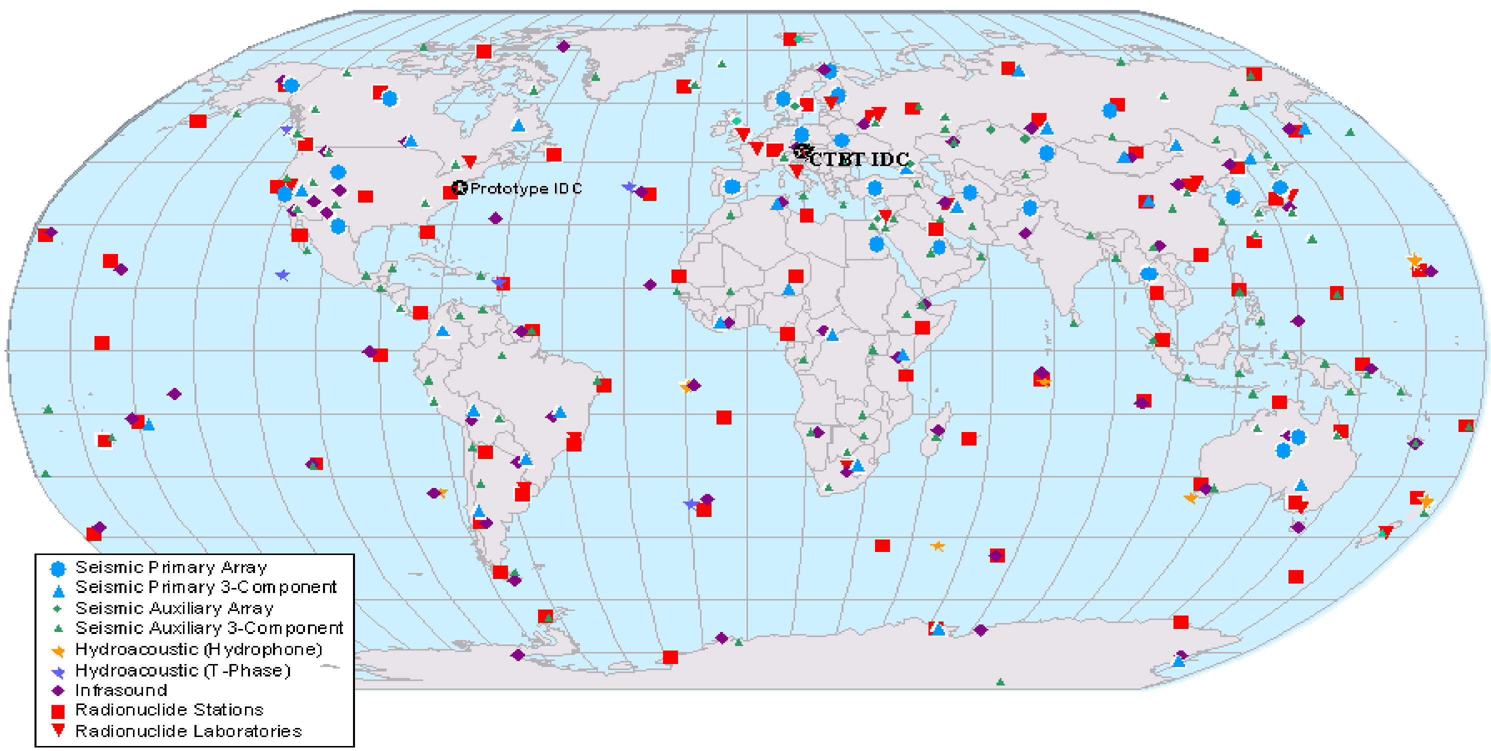

Figure 2. International Monitoring System. Final network as listed in the CTBT.

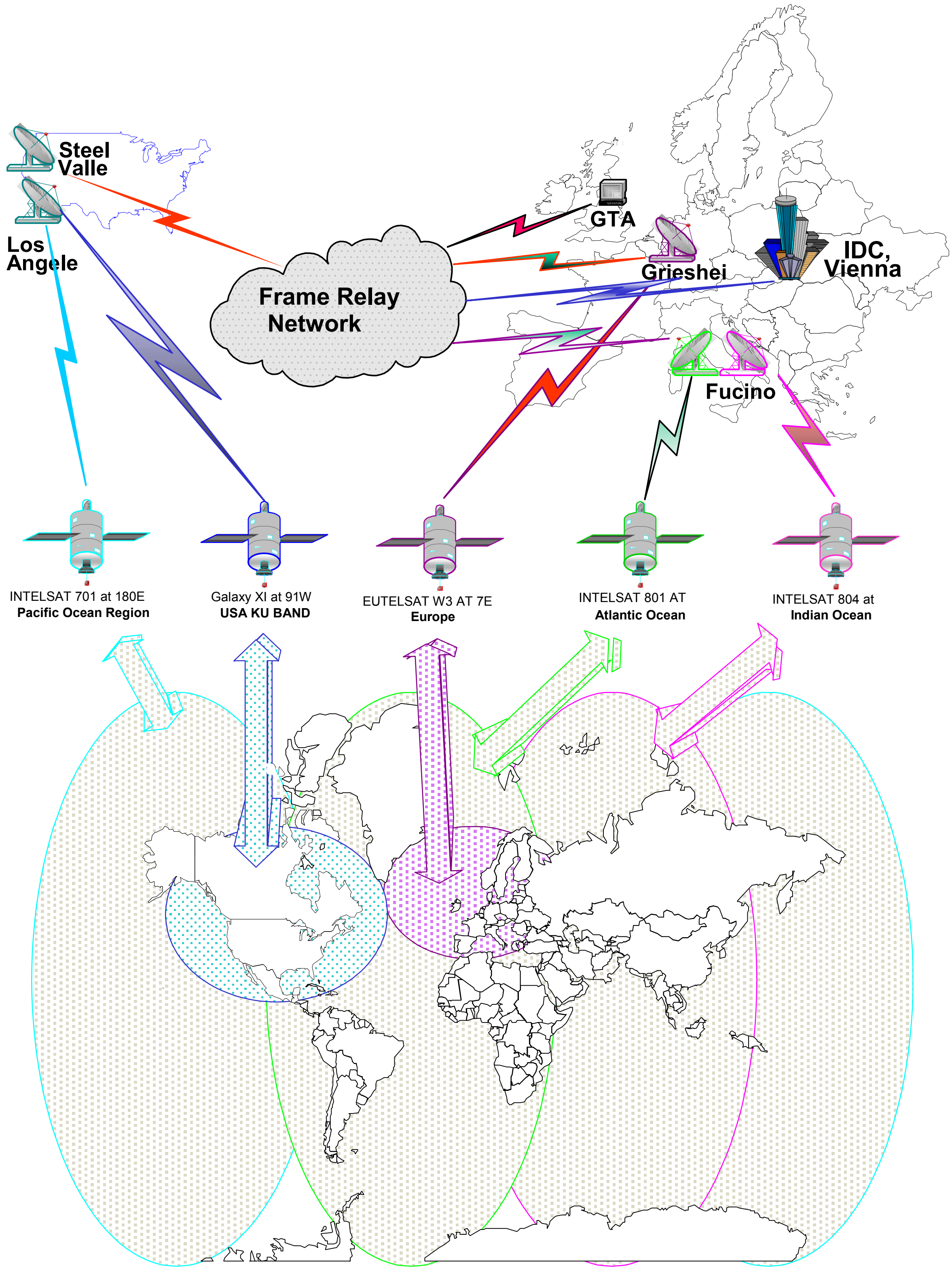

Figure 3. The central system of the Global Comunications Infrastructure (31 December) spans the entire globe (with a few exceptions in Antarctica) using space satellites covering the Pacific Ocean Region, Atlantic Ocean Region, Indian Region, Europe and North America. One hub is located in Griesheim, Germany and two each are located in Fucino, Italy and California, USA. The terrestrial links from the hubs to the IDC are secured by Frame Relay with ISDN backup. In addition there are terrestrial and satellite links to the IDC from five communications nodes operated by Member States.

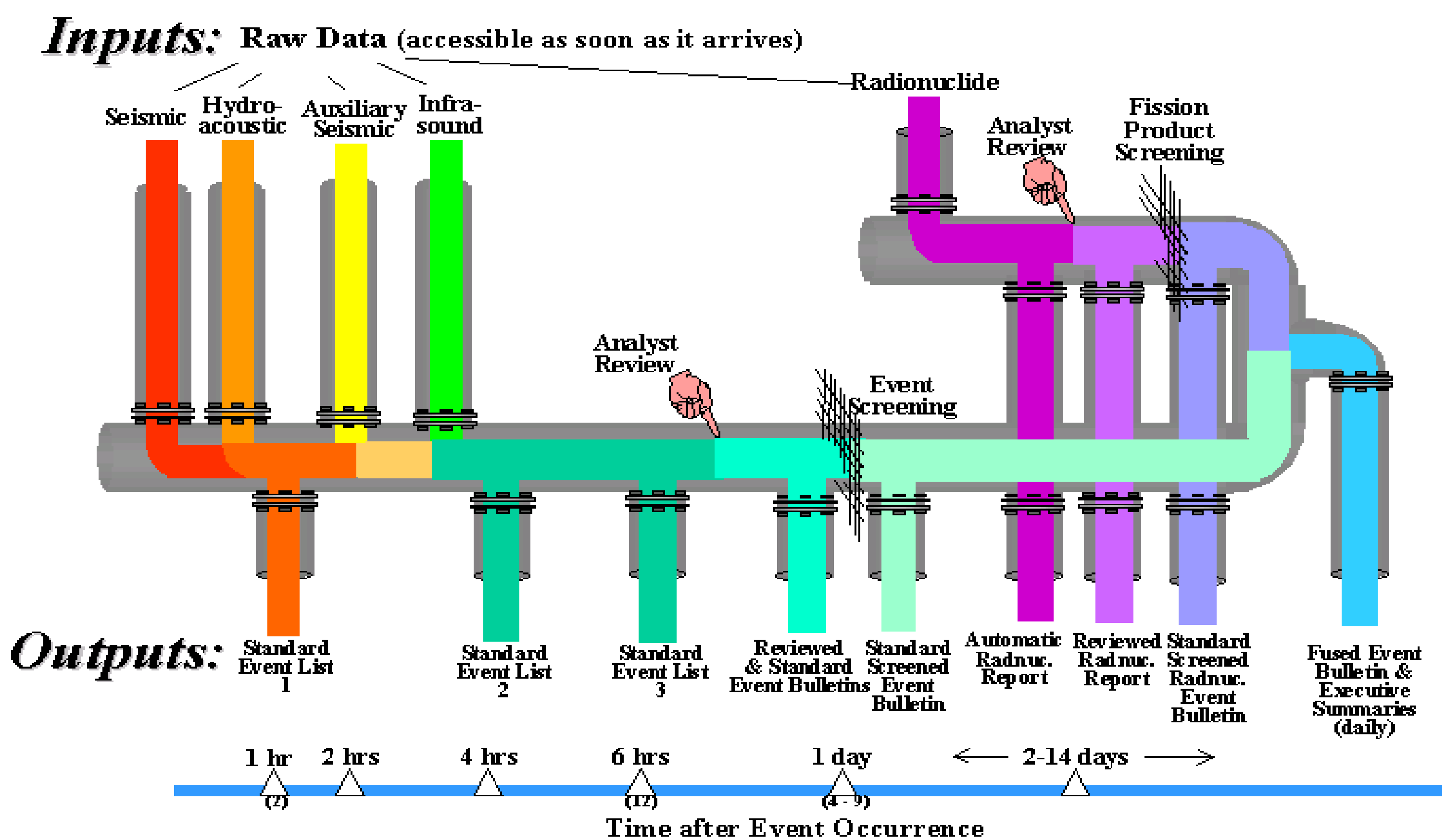

Figure 4. Stations of the International Monitoring System and several test stations that are transmitting data into the IDC for use in operational processing and analysis as of 31 December 2000.

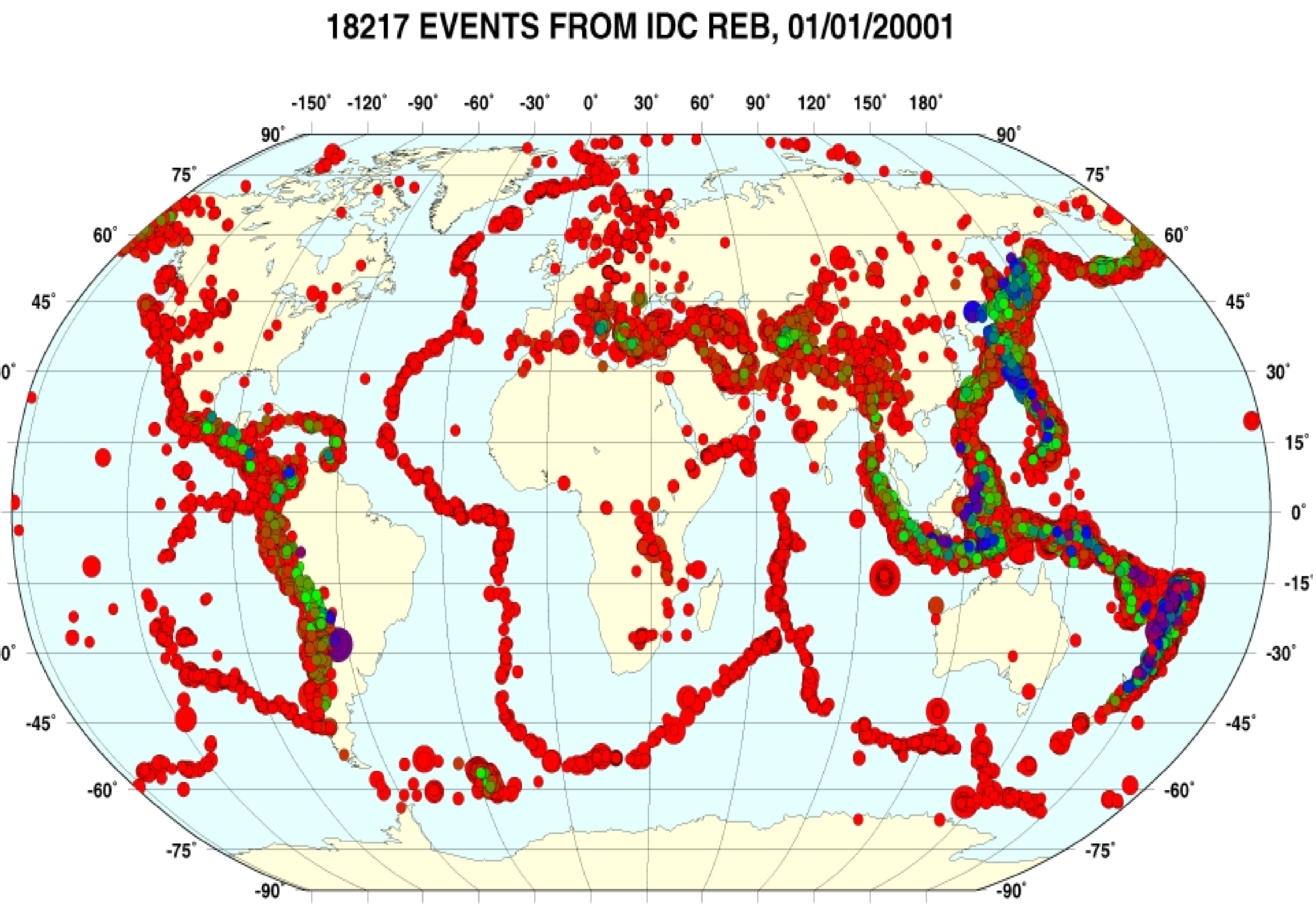

Figure 5. 18,21 7 Events from IDC Reviewed Event Bulletins: 1 January to 31 December 2000

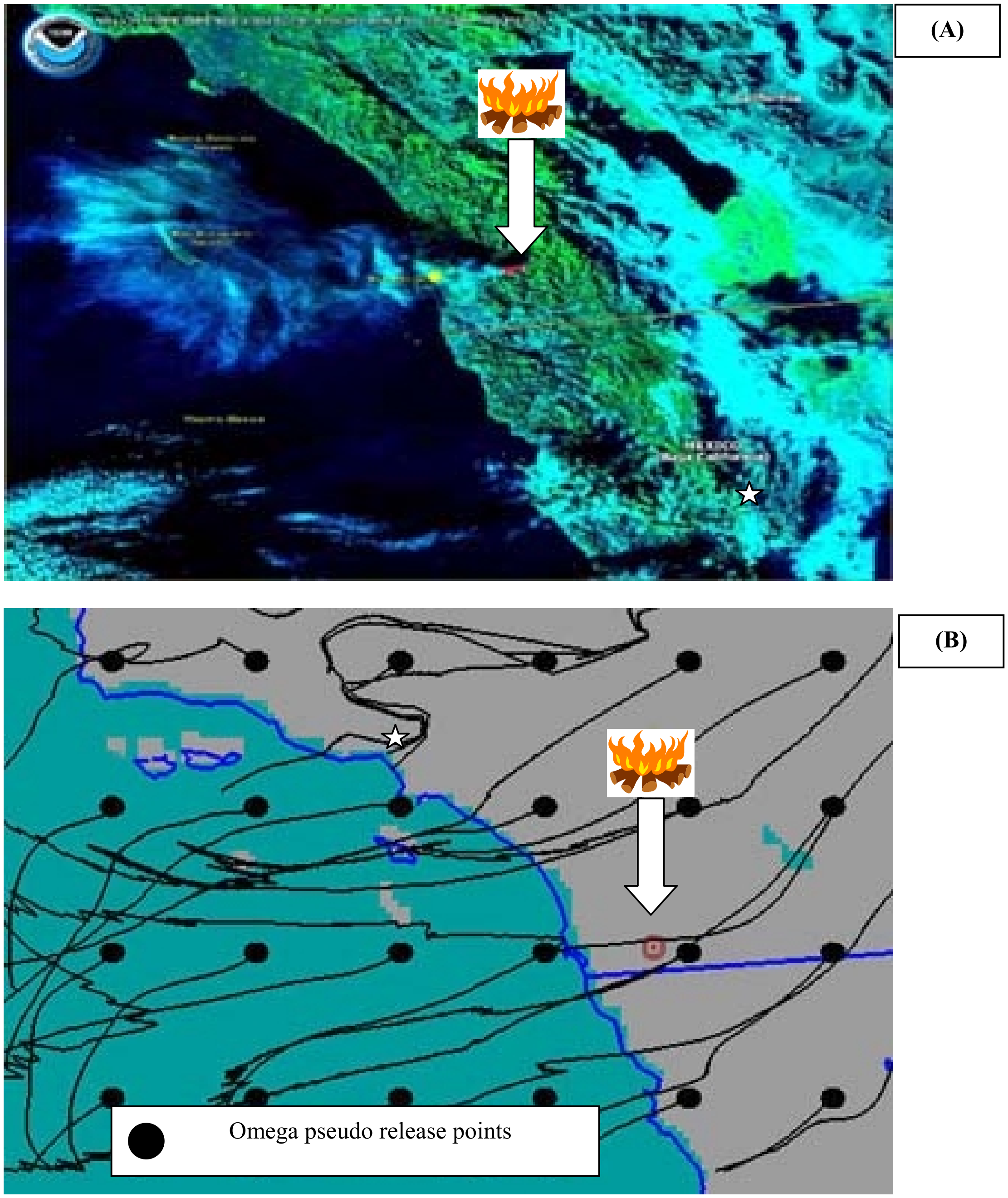

Figure 6: Smoke plume from a wild fire burning near San Diego, California in January 2001 as seen from space (A) and as simulated with the IDC atmospheric transport software (B). Shown in (B) are pseudo plumes released from predefined points on a regular grid. The simulation is done with the OMEGA [36] model. Note that the pseudo release particles are started from the release points at 000 UTC, while the fire started around 1500-1600 UTC.

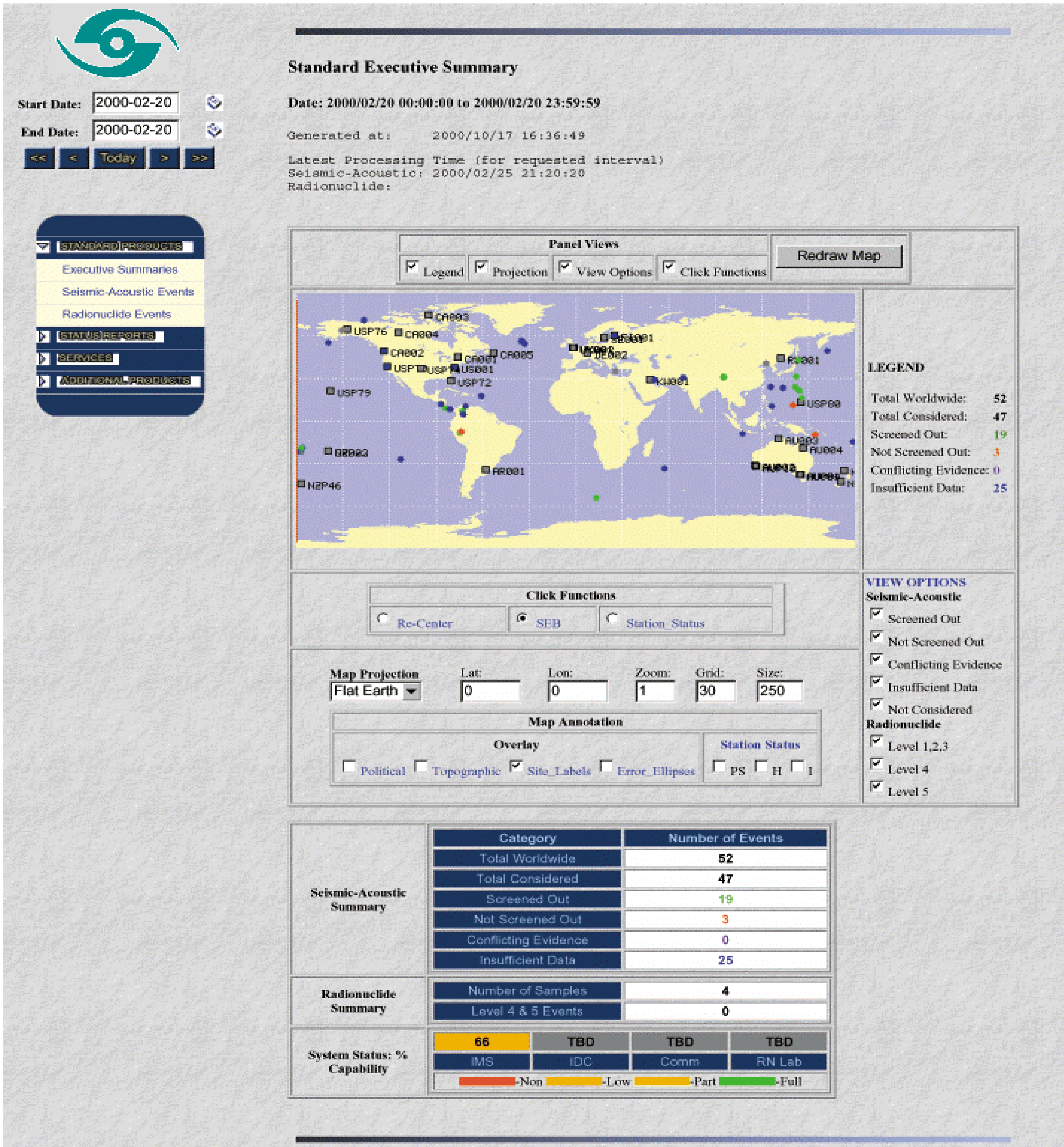

Figure 7. The Executive Summary page of the IDC secure Web site. This page highlights the events located and radionuclide detected for a particular day, the screening status of these, and also information on the state-of-health of the IMS, GCI and IDC systems. Users may select objects from this page, such as event locations from the map, to follow hyperlinks to many other pages with more detailed information, data and products. The left navigation bar also guides users to other data, products, services and documentation pages.

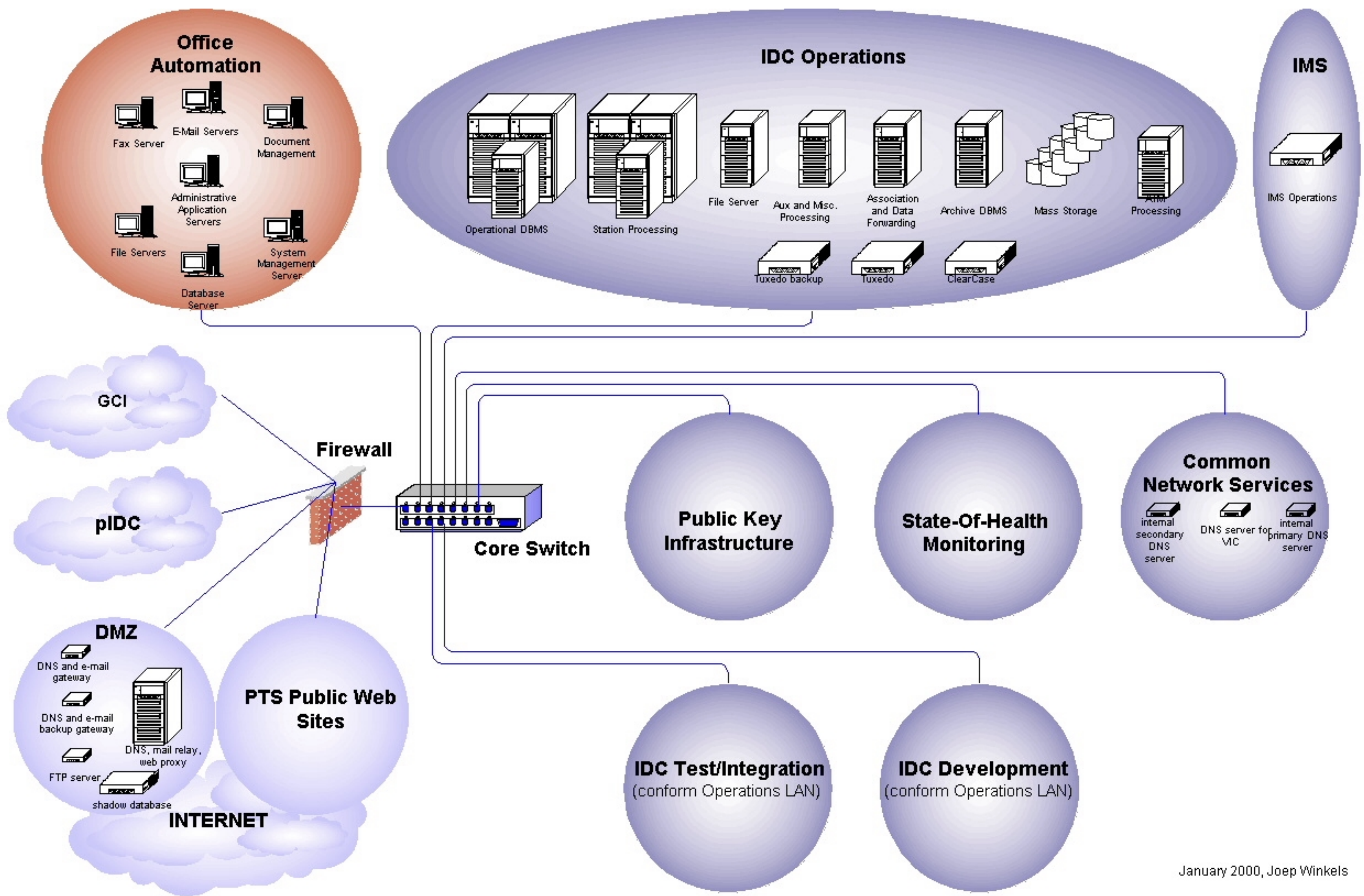

Figure 8. Illustrative view of the computer infrastructure of the IDC and the Provisional Technical Secretariat, showing the major Local Area Networks (LAN) and other services. The majority of the processes outlines in Section 3 are performed on the IDC Operations LAN, show at the top, middle of the figure.

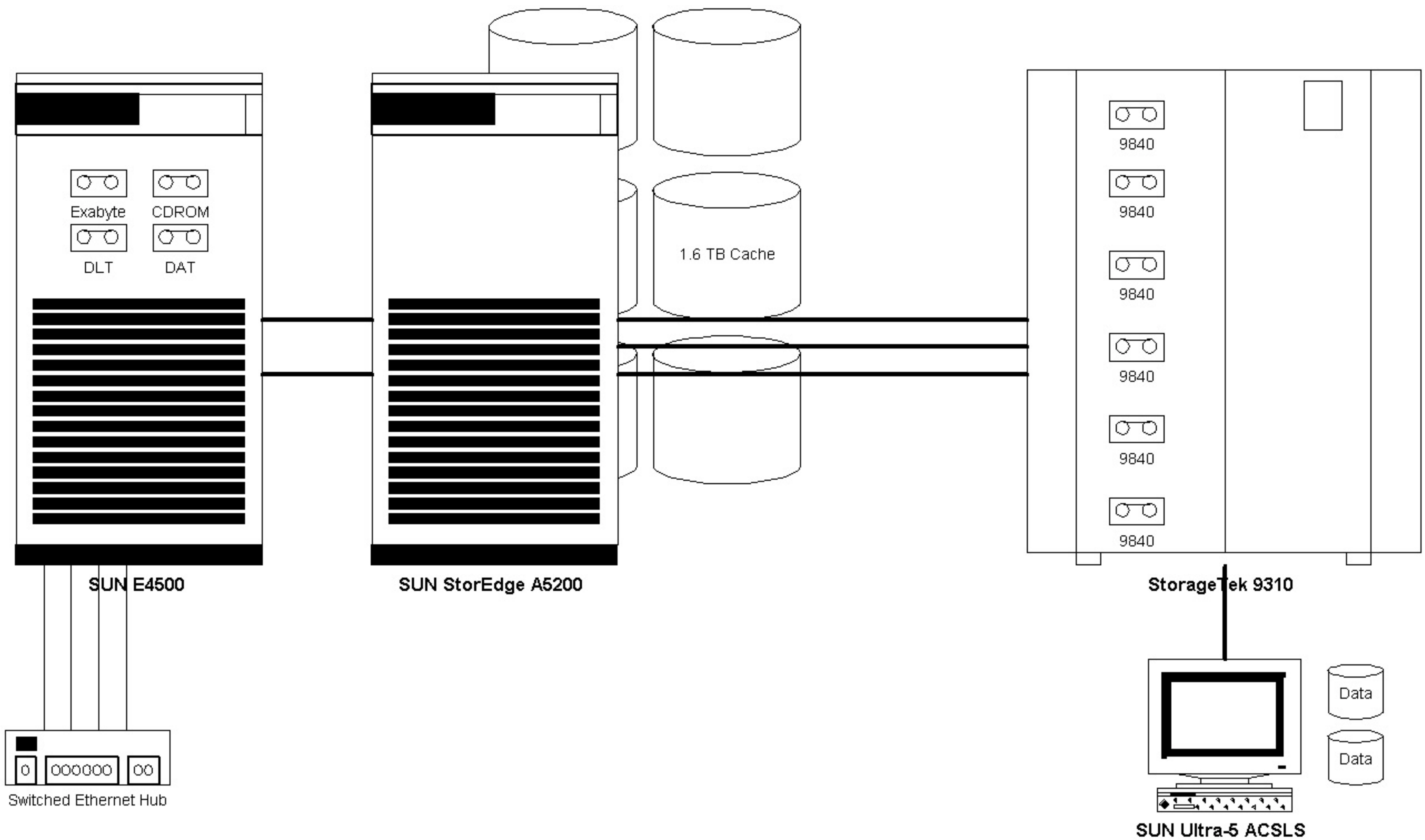

Figure 9. Architecture of the IDC Mass Storage System. See text for further details